

MATERIAL SUPLEMENTARIO

Tabla S1. Protocolo ODMAP (Resumen, Datos, Modelo, Evaluación, Predicción). Los principales pasos del modelado en el ciclo de modelado de distribución potencial de especies.

Table S1. ODMAP (Overview, Data, Model, Assessment, Prediction) protocol. The main modelling steps in the species potential distribution modelling cycle.

Elementos de ODMAP	Contenidos
INFORMACIÓN GENERAL	
Autoría	<ul style="list-style-type: none">• Autores: Melina Elizabeth Zuliani y Jorge Adrián Monjeau• Correo electrónico: melinazuliani@gmail.com• Título: Modelos de distribución potencial de mamíferos nativos en la Patagonia
Objetivo del modelo	<ul style="list-style-type: none">• Objetivo: mapeo/interpolación• Salidas del objetivo: mapas de probabilidad de presencia relativa
Taxa	15 especies de mamíferos nativos que se encuentran comúnmente en las comunidades de bosques y estepas del sur de la Argentina
Localización	Ecorregión del sur de la Argentina-Patagonia.
Escala de análisis	<ul style="list-style-type: none">• Extensión espacial (Lon/Lat): Longitud -60° O/-75° O, Latitud -30° S/-60° S.• Resolución espacial: 1 km• Extensión temporal/período de tiempo: los datos de ocurrencia se extienden desde la década de 1960 hasta el presente; los datos ambientales son del período 1970-2000• Tipo de límite de extensión: político
Resumen de datos de biodiversidad	<ul style="list-style-type: none">• Tipo de observación: ciencia ciudadana• Respuesta/tipo de datos: sólo presencia
Tipo de predictores	Climático, topográfico.
Modelo conceptual/Hipótesis	Hipótesis: la distribución de los mamíferos se encuentra relacionada al clima y la topografía
Supuestos	Asumimos que: <ul style="list-style-type: none">- Se incluyen los impulsores ecológicos relevantes (o sustitutos) de la distribución de especies- La detectabilidad no cambia a través de gradientes de hábitat- Las especies se encuentran en equilibrio con su entorno- El muestreo es adecuado y representativo (y cualquier sesgo se tiene en cuenta/se corrige)
Algoritmo	Técnica de modelo usado: MaxEnt. Justificación: elegimos el algoritmo de máxima entropía (MaxEnt) porque genera resultados óptimos utilizando sólo puntos de presencia, como es nuestro caso (Elith et al. 2006; Peterson et al. 2007). Además, se desempeña bien con tamaños de muestra pequeños (Wisz et al. 2008), incorpora efectos de interacción de variables ambientales y es un algoritmo determinista, lo que significa que los resultados siempre convergen en una distribución de probabilidad óptima única (Phillips et al. 2006). Es bastante robusto contra la autocorrelación espacial (Segurado et al. 2006), y con frecuencia supera los enfoques estadísticos tradicionales (Elith et al.

	2006)
	Complejidad del modelo: los modelos de ocupación solo incluyeron términos lineales (sin interacciones)
	Promedio del modelo: 10 repeticiones
Modelo workflow	<ul style="list-style-type: none"> - Antes de la construcción del modelo, todas las variables predictoras se estandarizaron y utilizaron el análisis de correlación de Pearson para evitar variables altamente correlacionadas - Sólo incluimos las variables más importantes y débilmente correlacionadas para cada especie - La contribución explicativa de cada variable se evaluó con análisis de <i>Jacknife</i> de la ganancia regularizada - El rendimiento del modelo predictivo se evaluó mediante una validación cruzada de diez veces - Realizamos un ajuste del parámetro de regularización (b) en modelos MaxEnt basados en AICc, utilizando el paquete ENMTools. Probamos los valores de b de 0.5 a 5, en pasos crecientes de 0.5.
Promedio de modelos/Conjuntos	<ul style="list-style-type: none"> - Generamos modelos preliminares para cada una de las especies usando tres multiplicadores de regularización (es decir, $k=0.5, 1$ y 2) (Warren and Seifert 2011; Radosavljevic and Anderson 2014), se utilizó el área bajo la curva (AUC) como un estimador del poder predictivo de cada modelo (Phillips et al. 2004, 2006). - Dentro de los modelos candidatos se seleccionaron según el valor mínimo para el AICc y promediados de acuerdo con sus correspondientes ponderaciones de Akaike (Bouchet and Meeuwig 2015). A continuación, se obtuvieron predicciones de consenso en todas las clases de modelos utilizando medias de conjuntos ponderadas
Software	<p>Software: MaxEnt (v3.4.0k)</p> <p>Code availability: https://biodiversityinformatics.amnh.org/open_source/maxent/</p> <p>Data availability: https://www.gbif.org/es/, https://sib.gob.ar/portada</p>

DATOS

Datos de biodiversidad	<p>Nombre del taxón: <i>Pudu puda</i> (pudú), <i>Hippocamelus bisulcus</i> (huemul), <i>Leopardus guigna</i> (gato huiña), <i>Leopardus geoffroyi</i> (gato montés), <i>Leopardus colocolo</i> (gato del pajonal), <i>Puma concolor</i> (puma), <i>Galictis cuja</i> (hurón), <i>Lyncodon patagonicus</i> (huroncito patagónico), <i>Conepatus chinga</i> (zorrino), <i>Lycalopex gymnocercus</i> (zorro gris), <i>Lycalopex culpaeus</i> (zorro colorado), <i>Chaetophractus villosus</i> (peludo), <i>Zaedyus pichiy</i> (piche), <i>Lama guanicoe</i> (guanaco) y <i>Dolichotis patagonum</i> (mara)</p> <p>Sistema de referencia taxonómica: http://cma.sarem.org.ar/</p> <p>Nivel ecológico: especie</p> <p>Fuente de datos: base de datos de adhesión, oportunista</p> <p>Tamaño de muestra: <i>Pudu puda</i> (76), <i>Hippocamelus bisulcus</i> (65), <i>Leopardus guigna</i> (14), <i>Leopardus geoffroyi</i> (27), <i>Leopardus colocolo</i> (27), <i>Puma concolor</i> (115), <i>Galictis cuja</i> (37), <i>Lyncodon patagonicus</i> (32), <i>Conepatus chinga</i> (30), <i>Lycalopex gymnocercus</i> (61), <i>Lycalopex culpaeus</i> (35), <i>Chaetophractus villosus</i> (46), <i>Zaedyus pichiy</i> (35), <i>Lama guanicoe</i> (75) y <i>Dolichotis patagonum</i> (41)</p> <p>Máscara regional: recortamos todos los datos al límite político de la Patagonia argentina</p> <p>Ajuste: se eliminaron los registros duplicados en las mismas localidades (por ejemplo, estaciones de llamada) y se minimizó la autocorrelación espacial eliminando aleatoriamente las ocurrencias dentro de los 10 km entre sí</p> <p>Limpieza: sólo se utilizaron las ocurrencias citadas como 'confirmadas' y con coordenadas geográficas para el registro de datos</p> <p>Datos de fondo: generamos 10000 puntos de fondo aleatorios dentro del área de estudio</p> <p>Errores y sesgos: errores potenciales de identificación y errores de georreferenciación</p>
Partición de datos	<p>Datos de entrenamiento: seleccionamos al azar el 75% de los datos para la calibración del modelo</p> <p>Datos de validación: 25% para la validación de las predicciones</p> <p>Datos de prueba: los modelos se evaluaron utilizando un método de validación cruzada de 10 veces</p>
Variables predictoras	Variables predictoras:

- Clima: utilizamos diferentes variables climáticas dependiendo de cada especie
 - Vegetación: dos variables que describen los cambios fenológicos anuales de la vegetación. Se utilizaron el índice de vegetación de diferencia normalizada (NDVI) y el índice de vegetación mejorado (EVI)
 - Terreno: altitud a una resolución vertical de 1 km
 - Perturbaciones: índice de huella humana (HFP)
- Fuente de datos:
- Clima: los datos climáticos se registraron de Worldclim 1.4 (Hijmans et al. 2005) y se prepararon como mapas ráster ASCII con una resolución de (0.93-0.93=0.86 km²), con el software QGIS (QGIS versión 3.16-Hannover)
 - Vegetación: el índice de vegetación de diferencia normalizada (NDVI) y el índice de vegetación mejorado (EVI) se obtuvieron del sensor satelital MODIS (resolución de 30 arcos, o ~ 1 km) (Carroll et al. 2004; Pettorelli et al. 2011)
 - Terreno: la variabilidad de la elevación se obtuvo del sensor basado en satélites MODIS (resolución de 30 arcos, o ~1 km)
 - Perturbación: se obtuvo el índice de huella humana (HFP) (Sanderson et al. 2002), del Centro de Aplicaciones y Datos Socioeconómicos (sedac)
- Extensión espacial: -180, -180, -60, -90 (xmin, xmax, ymin, ymax)
 Resolución espacial: la resolución primaria de los datos climáticos, NDVI, EVI, HFP y altitud tienen un tamaño de celda de una cuadrícula de 1 km
 Proyección geográfica: WGS 1984, EPSG:4326
 Extensión temporal:
- Clima: 1960-1990
 - Perturbación: 1995-2004
- Proceso de datos:
- Vegetación: se obtuvieron dos índices a partir del procesamiento de una serie de imágenes satelitales MODIS, las imágenes MOD13Q1 con fechas entre el 18 de febrero de 2000 y el 4 de marzo de 2000. Las imágenes consecutivas de 16 días de diferencia se combinaron utilizando el software Idrisi (Eastman 1990), lo que dio como resultado información de capa en una resolución espacial de 250 m

MODELO

Preselección de variables	Hipótesis de relevancia ecológica para la distribución de especies
Multicolinealidad	Colinealidad evaluada usando rango de Spearman. Enfoque de correlación cruzada, mediante la ejecución de tres modelos por especie (en cada escala), con diferentes subconjuntos de variables predictoras, basadas en diferentes niveles de dependencia de datos. Por lo tanto, definimos tres umbrales de correlación (de rango) entre las variables (con valores de rho de Spearman de 0.5, 1 y 2) para todas las especies. Usando estos umbrales definidos, generamos tres subconjuntos de predictores, a partir de los conjuntos originales. En cada caso se mantuvo el conjunto de variables no correlacionadas que mejor se ajustaba (por debajo del valor rho respectivo), descartándose las variables restantes
Configuración del modelo	Ajustes del modelo (ajuste) MaxEnt: no se proporcionan ajustes Configuración del modelo (extrapolación) MaxEnt: salida de registro, featureSet (Autofeatures: Linear, quadratic, product features, hinge features), regularizationMultiplierSet (b=1), convergenceThresholdSet (0.00001), samplingBiasRule (Índice de huella humana), Replicaciones (10), Porcentaje de prueba aleatorio (25), Prueba aleatoria (Sí), Tipo de ejecución de replicación (validación cruzada)
Estimación del modelo	Evaluación de los coeficientes del modelo: utilizamos el análisis <i>Jackknife</i> para identificar las variables que no contribuyen de forma importante a la solidez del modelo

EVALUACIÓN

Rendimiento en los datos de validación	Se utilizaron los valores promedios de ROC AUC como rendimiento predictivo del modelo sobre los datos de validación, siguiendo un procedimiento de 10-validación cruzada con diez repeticiones
Rendimiento en datos de prueba	Realizamos una validación cruzada y calculamos el área bajo la curva (AUC), eligiendo los valores promedios mayores a 0.7